

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-289122

(43)Date of publication of application : 27.10.1998

(51)Int.Cl.

G06F 11/20

G06F 12/00

(21)Application number : 09-110440

(71)Applicant : NEC CORP

(22)Date of filing : 11.04.1997

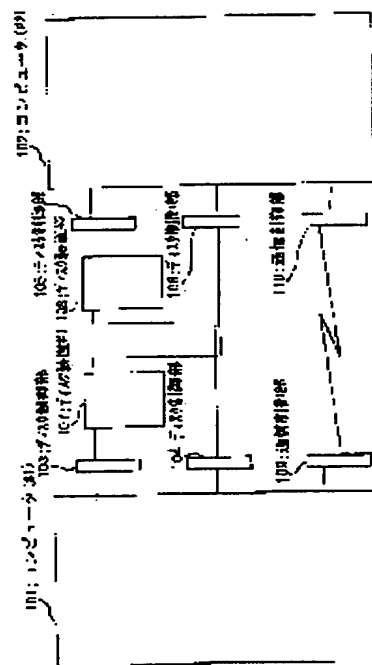
(72)Inventor : TAGUCHI TSUTOMU

(54) HIGH-SPEED CHANGEOVER SYSTEM FOR INTER-SYSTEM SHARED DISK IN HOT STANDBY SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To make a standby system quickly start a job without taking the consistency of data required in a file system or the like when a fault is generated in an active system and the job is to be switched to the standby system in the hot standby system.

SOLUTION: Two disks accessible from both of the active system and the standby system are used, and normally for the two disks, one (disk 1) is updated only from the active system and the other disk (disk 2) is updated only from the standby system. The two disks are logically used as one disk, and in the case of updating the data, updating is performed by the disk controller of the active system for the disk 1. For the disk 2, a request is issued so as to update the data to the standby system by using a communication route between the active system and the standby system, and in the standby system, the data of the disk 2 are updated by using a standby system disk controller only in the case that updating is requested from the active system. In the case that the fault is generated in the active system and the job is to be switched to the standby system, the standby system quickly starts the job by using the data of the disk 2 without the need of taking the consistency of the file system.



LEGAL STATUS

[Date of request for examination] 11.04.1997

[Date of sending the examiner's decision of rejection] 31.10.2000

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

BEST AVAILABLE COPY

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-289122

(43) 公開日 平成10年(1998)10月27日

(51) Int.Cl. ⁶	識別記号	F I
G 0 6 F 11/20	3 1 0	G 0 6 F 11/20 3 1 0 A
12/00	5 3 3	12/00 5 3 3 J

審査請求 有 請求項の数 4 F D (全 11 頁)

(21) 出願番号 特願平9-110440

(22) 出願日 平成9年(1997)4月11日

(71) 出願人 000004237

日本電気株式会社
東京都港区芝五丁目7番1号

(72) 発明者 田口 勉

東京都港区芝五丁目7番1号 日本電気株
式会社内

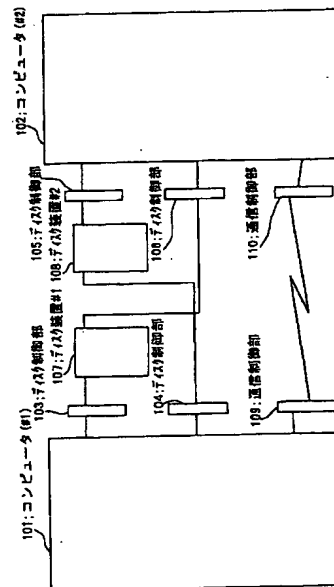
(74) 代理人 弁理士 加藤 朝道

(54) 【発明の名称】 ホットスタンバイシステムにおける系間共用ディスクの高速切り替え方式

(57) 【要約】

【課題】 ホットスタンバイシステムにおいて、現用系に障害が発生し待機系に業務を切り替える時に待機系がファイルシステムなどで必要となるデータの整合性を取ることなく迅速に業務を開始する。

【解決手段】 現用系／待機系の両系からアクセス可能なディスクを2台使用し、通常この2台のディスクは、1台（ディスク1）は現用系からのみ更新し、もう一方のディスク（ディスク2）は待機系からのみ更新する。2台のディスクは論理的に1台のディスクとして使用され、データを更新する場合、ディスク1に対しては現用系のディスクコントローラにより更新し、ディスク2に対しては現用系-待機系間の通信経路を使用して待機系にデータを更新するように要求を発行し、待機系では、現用系から更新要求があった場合のみ待機系ディスクコントローラを使ってディスク2のデータを更新する。現用系に障害が発生し待機系に業務を切り替える場合、待機系はファイルシステムの整合性を取る必要がなくディスク2のデータを使い迅速に業務を開始することができる。



【特許請求の範囲】

【請求項1】現用系及び待機系の両系からアクセス可能な2系統のディスク装置を備え、

現用系及び待機間で共有されるデータは2系統のディスク装置上に格納され、

通常運用時において、共有データを更新する場合、現用系が第1系のディスク装置に対して自系のディスク制御装置を通して更新し、第2系のディスク装置に対しては、現用系が、現用系と待機系間の通信経路を介して待機系に対してデータの更新要求を発行し、待機系は、現用系から更新要求を受けて自系のディスク制御装置を介して前記第2系のディスク装置のデータの更新を行い、現用系に障害が発生し待機系に業務を切り替える場合は、待機系は、前記第2系のディスク装置のデータを用いて業務を開始する、ようにしたことを特徴とする、ホットスタンバイシステムにおける系間共用ディスクの高速切り替え方式。

【請求項2】待機系に業務切り替えた後は、待機系が、前記第2のディスク装置から前記第1のディスク装置へデータを複写して、前記第2のディスク装置と前記第1のディスク装置の内容を一致させ、データを更新する場合、待機系のディスク制御装置を通して前記第1のディスク装置と前記第2のディスク装置のデータを更新する、ようにしたことを特徴とする、請求項1記載のホットスタンバイシステムにおける系間共用ディスクの高速切り替え方式。

【請求項3】業務を実行している現用系のコンピュータと、

現用系のコンピュータの障害発生時にその業務を引き継ぐ待機系のコンピュータを備え、

現用系/待機系の両系からアクセス可能なディスク装置上に系間で共有するデータを格納しているホットスタンバイシステムにおいて、

両系からアクセス可能な2台のディスク装置と、通常運用時は、2台のディスク装置のうち第1のディスク装置は現用系からのみ更新し、第2のディスク装置は待機系からのみ更新し、

共有ディスク上のデータを、現用系から更新する場合には、前記第1のディスク装置に対しては、現用系のディスク制御装置により更新し、前記第2のディスク装置に対しては現用系と待機系間の通信経路を使用して待機系にデータを更新するように要求を発行し、待機系では、現用系から更新要求があった場合のみ待機系のディスク制御装置を用いて前記第2のディスク装置のデータを更新し、前記第1、第2のディスク装置上に同一のデータを格納し、論理的に1台のディスク装置として使用する手段と、

現用系に障害が発生し待機系に業務を切り替える場合には、待機系が、前記第2のディスク装置上のデータを整合性をとることなく、前記第2のディスク装置上のデータ

タを用いて業務を開始する、

ことを特徴とする、ホットスタンバイシステムにおける系間共用ディスクの高速切り替え方式。

【請求項4】待機系に業務切り替えた後は、前記第2のディスク装置と前記第1のディスク装置の内容を一致させるために、待機系が前記第2のディスク装置から前記第1のディスク装置へデータをコピーし、

データを更新する場合には、待機系のディスク制御装置を通して前記第1のディスク装置と前記第2のディスク装置のデータを更新する手段と、

現用系が復帰した場合には、前記第1のディスク装置を現用系のディスク制御装置により更新するように設定変更する、

ことを特徴とする、請求項3記載のホットスタンバイシステムにおける系間共用ディスクの高速切り替え方式。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、コンピュータの二重化システムに関し、特に、業務を実行している現用系コンピュータと現用系コンピュータの障害発生時にその業務を引き継ぐ待機系のコンピュータで二重化構成のシステムを構成し、システムの信頼性を向上させるコンピュータシステムに関する。

【0002】

【従来の技術】この種の二重化構成システムの従来の技術として、例えば特開平6-175788号公報には、ディスク装置へのアクセスの競合を少なくして性能の向上を図り、現用機で障害が発生しても実行途中のプロセスを引き継ぐことのできるバックアップ装置として、図11に示すようなシステム構成が提案されている。図11において、11は現用系、17は待機系を示しており、4はディスクサブシステム、5は共有ディスク装置である。

【0003】図11を参照して、現用系11と待機系17はそれぞれ内蔵ディスク装置11-7、17-7を所有し、また共有ディスク装置5を設け、内蔵ディスク装置と同一の内容を書き込むことにより二重化構成とし、現用機は、通常、アクセス時間の短い内蔵ディスク装置からデータを読み込み、自身の内蔵ディスク装置11-7と共有ディスク装置5に共有データを書き込み、現用機で障害が発生した場合は、待機系は共有ディスク装置5から引継情報を読み出すことにより、現用系の処理を引き継ぎ、待機系は共有ディスク装置5の内容を自身の内蔵ディスク装置17-7にコピーするようにしている。

【0004】

【発明が解決しようとする課題】業務を実行しているコンピュータ（現用系）と現用系のコンピュータの障害発生時にその業務を引き継ぐコンピュータ（待機系）で構成されるホットスタンバイシステムにおいて、両系間か

らアクセス可能な共有ディスク装置上に、バッファキャッシュを使用して、高速にアクセス可能なファイルシステムを構築し、そのファイルシステム上に業務で使用するデータを格納している場合に、現用系に障害が発生したため、待機系に業務を切り替えた場合、共有ディスク装置上のデータを待機系で使用するためには、バッファキャッシュを使用しているためファイルシステムの整合性をとってから業務を開始しなければならない。

【0005】このため、ファイルシステムに格納されているデータ量が多い場合、整合性をとるのに時間を要することになり、現用系から待機系への業務の切り替えを遅延させている。

【0006】したがって、本発明は、上記問題点を鑑みてなされたものであって、その目的は、ホットスタンバイシステムにおいて、現用系に障害が発生し待機系に業務を切り替える時に、待機系がファイルシステムなどで必要となるデータの整合性を取ることなく、迅速に業務を開始することを可能とした、ホットスタンバイシステムにおける系間共用ディスクの高速切り替え方式を提供することにある。

【0007】

【課題を解決するための手段】前記目的を達成するため、本発明は、現用系及び待機系の両系からアクセス可能な2系統のディスク装置を備え、現用系及び待機間で共有されるデータは2系統のディスク装置上に格納され、通常運用時において、共有データを更新する場合、現用系が第1系のディスク装置に対して自系のディスク制御装置を通して更新し、第2系のディスク装置に対しては、現用系が、現用系と待機系間の通信経路を介して待機系に対してデータの更新要求を発行し、待機系は、現用系から更新要求を受けて自系のディスク制御装置を介して前記第2系のディスク装置のデータの更新を行い、現用系に障害が発生し待機系に業務を切り替える場合は、待機系は、前記第2系のディスク装置のデータを用いて業務を開始する、ようにしたことを特徴とする。

【0008】また、本発明においては、待機系に業務切り替えた後は、待機系が、前記第2のディスク装置から前記第1のディスク装置へデータを複写して、前記第2のディスク装置と前記第1のディスク装置の内容を一致させ、データを更新する場合、待機系のディスク制御装置を通して前記第1のディスク装置と前記第2のディスク装置のデータを更新する、ようにしたことを特徴とする。

【0009】

【発明の実施の形態】本発明の実施の形態について以下に説明する。本発明は、その好ましい実施の形態として、ホットスタンバイシステムにおいて、現用系／待機系の両系からアクセス可能な第1、第2のディスク装置（図1の107、108）を備え、第1のディスク装置は現用系からのみ更新し、第2のディスク装置は待機系

からのみ更新し、これら2台のディスク装置は論理的に1台のディスク装置として使用する。そして、データを更新する場合には、第1のディスク装置（図1の107）に対しては現用系（図1の101）のディスク制御部（図1の103）により更新し、第2のディスク装置（図1の108）に対しては通信制御部（図1の109）を使用して待機系（図1の102）にデータを更新するように要求を発行する。待機系（図1の102）では、現用系から更新要求があった場合にのみ、自系のディスク制御部（図1の105）を使って第2のディスク装置（図1の108）のデータを更新する。

【0010】このため、現用系に障害が発生し待機系に業務を切り替える場合は、待機系はファイルシステムの整合性を取る必要がなく、第2のディスク装置のデータを使って迅速に業務を開始することができる。

【0011】本発明は、その好ましい実施の形態において、現用系に障害が発生し待機系に業務切り替えた後は、第2のディスク装置（図1の108）と第1のディスク装置（図1の107）の内容を一致させるために、待機系が第2のディスク装置から第1のディスク装置へデータをコピーし、データを更新する場合には、待機系のディスク制御装置（図1の106、105）を通して第1のディスク装置（図1の107）と第2のディスク装置（図1の108）のデータを更新する。

【0012】そして、現用系が障害から復帰した場合には、第1のディスク装置（図1の107）を現用系のディスク制御装置（図1の103）により更新するように変更する。

【0013】

【実施例】上記した本発明の実施の形態について更に詳細に説明すべく、本発明の実施例を図面を参照して以下に説明する。図1は、本発明の一実施例の構成を示すブロック図である。

【0014】まず図1を参照して、本実施例のシステムのハードウェア構成を説明する。本実施例は、現用系として運用されるコンピュータ（#1）101と、待機系として運用され、コンピュータ101に障害が発生した場合には現用系に切り替えられるコンピュータ（#2）102と、コンピュータ101とコンピュータ102の2台から制御可能であり、業務で使用するデータを格納したディスク装置107とディスク装置108と、ディスク装置107をコンピュータ101から制御するためのディスク制御部（ディスクコントローラ）103と、ディスク装置108をコンピュータ101から制御するためのディスク制御部104と、ディスク装置108をコンピュータ102から制御するためのディスク制御部105と、ディスク装置107をコンピュータ102から制御するためのディスク制御部106と、コンピュータ101とコンピュータ102間でデータを通信するための通信制御部109、110と、を備えて構成さ

れている。

【0015】図2を参照して、本実施例のソフトウェア構成について説明する。ソフトウェアは、コンピュータ（#1）101、コンピュータ（#2）102上で同じものが動作しているものとする。

【0016】業務を実行する業務プログラム201と、データアクセスプログラム202と、データアクセス要求通信プログラム203と、系切り替えプログラム204と、ディスク間データ整合プログラム205と、自系が現用系か待機系であるかを示す#1系状態管理領域206と、他系が現用系か待機系であるかを示す#2系状態管理領域207と、を備えている。

【0017】図2において、208は業務プログラム201から発行されるデータアクセス要求、209はデータアクセスプログラム202が発行する2台のディスクコントローラに対するデータアクセス要求、210はデータアクセスプログラム202がデータアクセス要求通信プログラム203が発行する他系へのデータアクセス要求、をそれぞれ示している。

【0018】また、211はデータアクセス要求通信プログラム203が自系/他系の状態を確認するための自系状態管理領域206と他系状態管理領域207の参照、213はデータアクセス要求通信プログラム203が通信制御部に発行する他系へのデータアクセス要求の送信要求、212はデータアクセス要求通信プログラム203が通信制御部から受け取る他系からのデータアクセス要求の受信、214はデータアクセス要求通信プログラム203がデータアクセスプログラム202に発行するデータアクセス要求、をそれぞれ示している。

【0019】そして、215はデータアクセスプログラム202が自系/他系の状態を確認するための自系状態管理領域206と他系状態管理領域207の参照、216は系切り替え時、相手系復旧時に系切り替えプログラム204が行う自系状態管理領域206と他系状態管理領域207を更新、217は系切り替え時、相手系復旧時に系切り替えプログラム204がディスク間データ整合プログラム205に対して発行するデータ整合要求、218はディスク間データ整合プログラム205がデータアクセスプログラム202に対して発行するディスク間のデータの整合性を取るためのデータアクセス要求、219は系切り替えプログラム204がデータアクセスプログラム202に対して発行する他系復旧処理要求、をそれぞれ示している。

【0020】次に、本実施例の動作について説明する。図3に、通常運用時（現用系、待機系とも運用中）のコンピュータ（#1）101におけるデータ更新処理をフローチャートとして示す。

【0021】データアクセスプログラム202は、データ更新要求を受けると、#1系状態管理領域206を参照し、自系が現用系、待機系のいずれであるかを調べる

（ステップ3-1）。

【0022】自系が現用系であった場合には、ディスク制御部103に対しディスク装置107の更新要求を発行し、ディスク装置107の内容を更新する（ステップ3-2）。一方、自系が待機系である場合には、後に説明する待機系データ更新処理を実行する。

【0023】自系が現用系であった場合には、次に、#2系状態管理領域207を参照し、相手系である#2系が運用中であるか、あるいはダウン中であるかを調べる（ステップ3-3）。

【0024】相手系の状態が運用中であれば（ステップ3-3のNo分岐）、データアクセス要求通信プログラム203に対しコンピュータ（#2）102に、ディスク装置（#1）107の更新と同じアクセス要求を送信するように要求を発行する（ステップ3-4）。データアクセス要求通信プログラム203は、データアクセス要求の送信要求を受けると、通信制御部109にその要求を発行する（図5のステップ5-1）。

【0025】一方、相手系がダウン中であれば（ステップ3-3のYes分岐）、自系のディスク制御部104に対し、ディスク装置（#2）108の更新要求を発行し、ディスク装置#2の内容を更新する（ステップ3-5）。

【0026】以上により、現用系の更新要求は、ディスク装置（#1）107とディスク装置（#2）108に対して行われる。

【0027】図6は、本実施例において、更新要求を受信した時の、データアクセス要求通信プログラム203の動作を説明するためのフローチャートである。コンピュータ（#1）101からデータアクセス要求を受信したコンピュータ（#2）102上のデータアクセス要求通信プログラム212は、コンピュータ上102の#2系状態管理領域207を参照して自系が現用系、待機系のいずれであるかを調べる（ステップ6-1）。

【0028】自系が待機系である場合のみ、データアクセスプログラム202に対しディスク装置（#2）108に対する更新要求を発行する（ステップ6-2）。

【0029】図4は、待機系におけるデータ更新処理の動作を説明するためのフローチャートである。

【0030】図4を参照して、待機系のデータアクセスプログラム202がデータ更新要求を受けると、要求がデータアクセス要求通信プログラム203から発行されたものであるか否かを調べる（ステップ4-1）。そして、データアクセス要求通信プログラムからの要求であった場合のみ（ステップ4-1のYes分岐）、ディスク装置（#2）108のデータを更新する（ステップ4-2）。

【0031】以上により、待機系では、現用系からの更新要求以外ディスク更新を処理しない制御を実行する。通常運用中は以上のような処理を実行している。

【0032】以下に、本実施例において、現用系が障害発生等でダウンし、系切り替えが発生し、待機系が現用系に切り替わった時点から、ダウンしたシステムが復旧し、待機系として運用を開始するまでの処理を示す。

【0033】図7は、本実施例において、待機系の系切り替えプログラム204が現用系のダウンを検出した時の処理をフローチャートで示したものである。コンピュータ(2)102が待機系である時に、コンピュータ(2)102上の系切り替えプログラム204は、コンピュータ101のダウンを検出すると、#2系状態管理領域207の内容を待機系から現用系に変更する(ステップ7-1)。

【0034】さらに#1系状態管理領域208の内容を、運用中から、ダウン中に変更する(ステップ7-2)。その後、ディスク間データ整合要求をディスク間データ整合プログラム206に発行する(ステップ7-3)。

【0035】ディスク間データ整合プログラム205は、このディスク間データ整合要求を受け取ると、ディスク装置(2)108の全内容を読み込み、読み込んだデータをディスク装置(1)107に書き込み、ディスク装置(1)107と108ディスク装置(2)108の内容を一致させる(図8のステップ8-1)。

【0036】系切り替え後、コンピュータ(2)102からの更新要求は、図3のディスク制御部1とディスクコントローラ2、ディスク装置1とディスク装置2を逆にして、ステップ3-1、3-2、3-3、3-5の処理を実行している。ディスク装置(2)108は、切り替え前も、コンピュータ(2)102から更新されているため、バッファキャッシュの同期は必要がない。

【0037】図9は、本実施例において、系切り替えプログラム204が相手系のダウン検出後に、相手系の復旧を検出した場合の処理を示すフローチャートである。

【0038】系切り替えプログラム204は、図8で示したデータ整合処理が完了したか調べる(ステップ9-1)。完了していない場合には、一定時間待ち(ステップ9-3)、データ整合処理が完了したか調べる。

【0039】データ整合処理が完了するまでは、ステップ9-1、9-3の処理を繰り返す。

【0040】データ整合処理が完了すると、他系状態管理領域の状態を運用中に変更し(ステップ9-2)、データアクセスプログラムに他系復旧処理要求を発行する(ステップ9-4)。

【0041】図10は、本実施例において、データアクセスプログラム202が他系復旧処理要求を受け取った時の処理を示すフローチャートである。データアクセスプログラム202は他系復旧処理要求を受け取るとディスク装置(1)107とディスク装置(2)108

へのアクセス処理を一旦停止する(ステップ10-1)。

【0042】その後、ディスク装置(1)107をコンピュータ(1)101の制御下に移すため、ディスク装置(1)のバッファキャッシュの内容をディスク装置(1)107に反映する(ステップ10-2)。

【0043】最後に、ディスク装置(2)108へのアクセス処理を再開し、ディスク装置(1)107に対する要求は、データアクセス要求通信プログラム203に送信要求を発行し、コンピュータ(1)101を待機系、コンピュータ(2)108を現用系としての通常運用に移行する(ステップ10-3)。

【0044】

【発明の効果】以上説明したように、本発明によれば、ホットスタンバイシステムにおいて、現用系に障害が発生し待機系に業務を切り替える場合は、待機系はファイルシステムの整合性を取る必要がなく、迅速に業務を開始することができる、という効果を奏する。

【0045】その理由は、本発明においては、現用系/待機系の両系からアクセス可能なディスクを2台使用し、通常この2台のディスク装置を、1台は現用系からのみ更新し他方のディスク装置は、現用系-待機系間の通信経路を使用して待機系にデータを更新するように要求を発行して更新し、待機系では、現用系から更新要求があった場合のみ待機系ディスク制御装置を介してディスク装置のデータを同じ内容で更新し、現用系に障害が発生し待機系に業務を切替る場合に、待機系は、待機系を介して更新をしていたディスク装置のデータを用いて迅速に業務を開始することができる。

【図面の簡単な説明】

【図1】本発明の一実施例のハードウェアの構成を示す図である。

【図2】本発明の一実施例のソフトウェアの構成を示す図である。

【図3】本発明の一実施例において、通常運用時(現用系、待機系とも運用中)のコンピュータにおけるデータ更新処理を示す流れ図である。

【図4】本発明の一実施例において、待機系におけるデータ更新処理の動作を説明するための流れ図である。

【図5】本発明の一実施例において、データ更新要求送信処理を示す流れ図である。

【図6】本発明の一実施例において、更新要求を受信した時の、データアクセス要求通信プログラムの動作を説明するための流れ図である。

【図7】本発明の一実施例において、待機系の系切り替えプログラムが現用系のダウンを検出した時の処理を示す。

【図8】本発明の一実施例において、ディスク間データ整合処理を示す流れ図である。

【図9】本発明の一実施例において、系切り替えプロ

ラムが相手系のダウン検出後に、相手系の復旧を検出した場合の処理を示す流れ図である。

【図10】本発明の一実施例において、データアクセスプログラムが他系復旧処理要求を受け取った時の処理を示す流れ図である。

【図11】従来の技術を説明するための図である。

【符号の説明】

101 コンピュータ（#1）

102 コンピュータ（#2）

103 ディスク制御部（ディスクコントローラ）

104 ディスク制御部

* 105、106 ディスク制御部

107、108 ディスク装置

109、110 通信制御部

201 業務を実行する業務プログラム

202 データアクセスプログラム

203 データアクセス要求通信プログラム

204 系切り替えプログラム

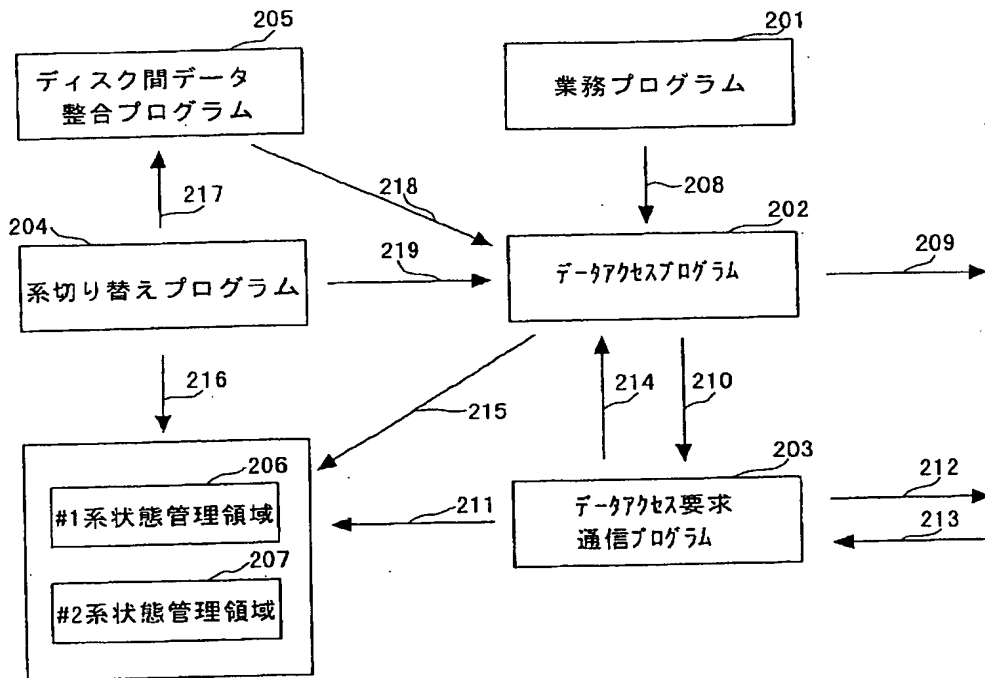
205 ディスク間データ整合プログラム

206 #1系状態管理領域

207 #2系状態管理領域

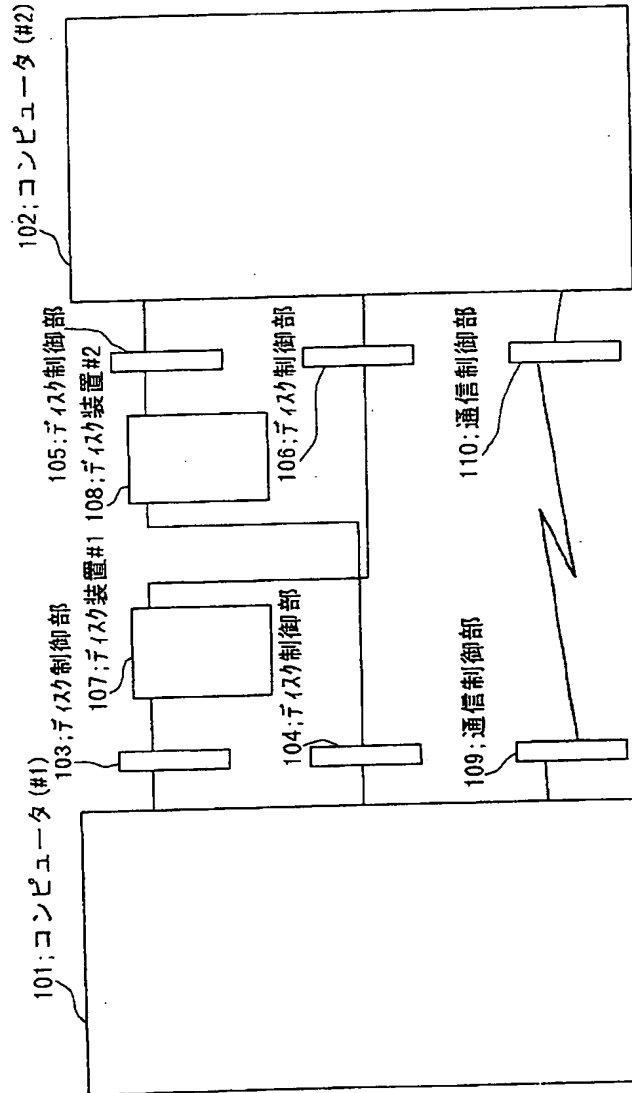
*

【図2】

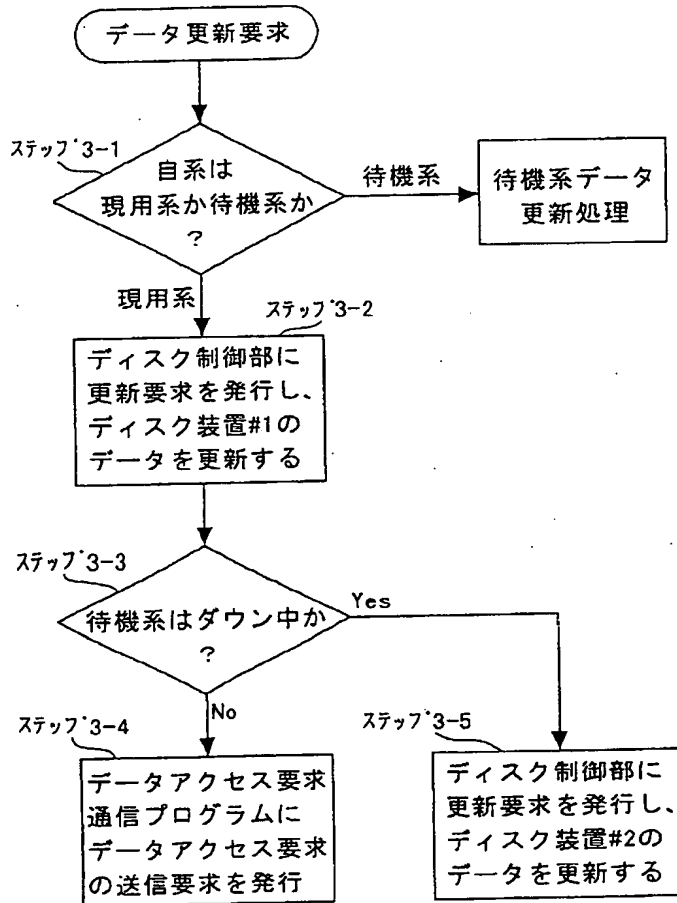


(7)

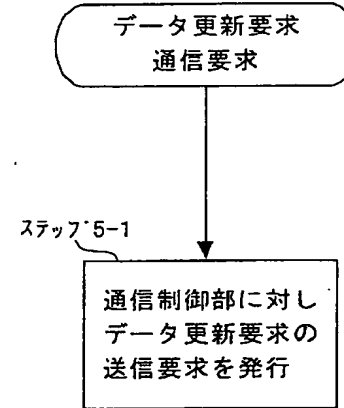
【図1】



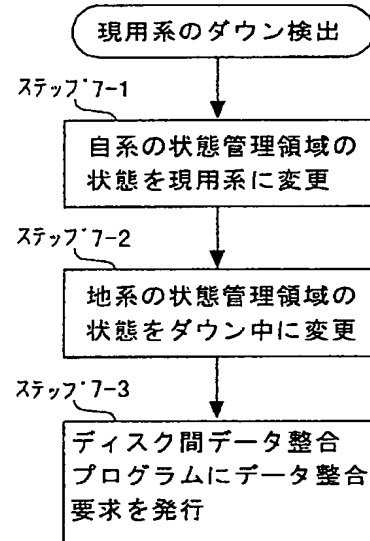
【図3】



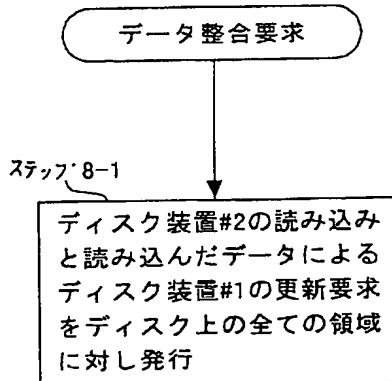
【図5】



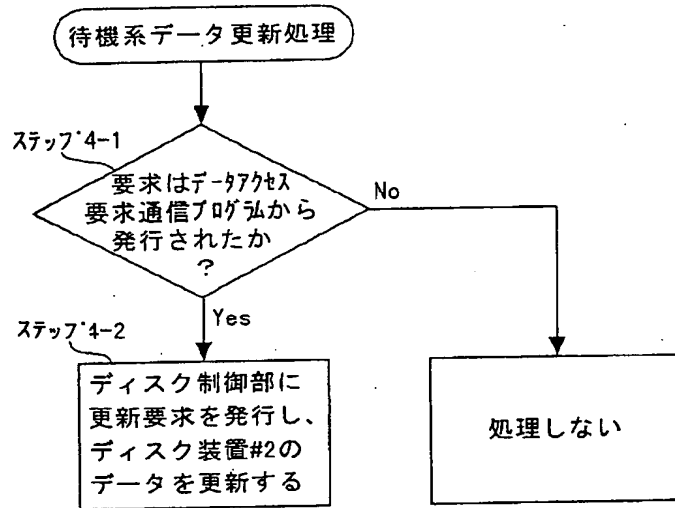
【図7】



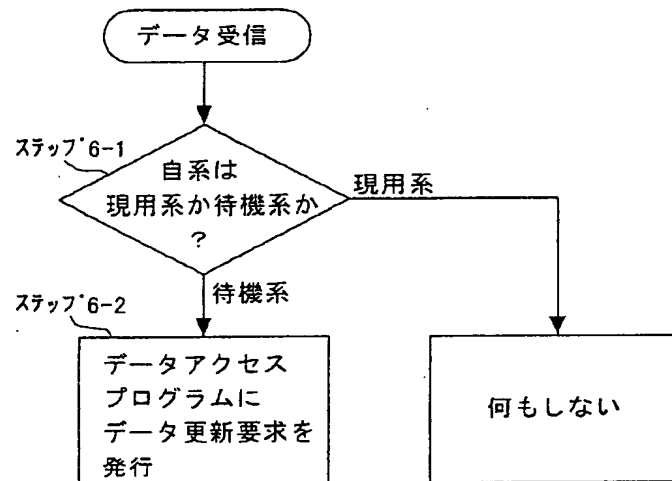
【図8】



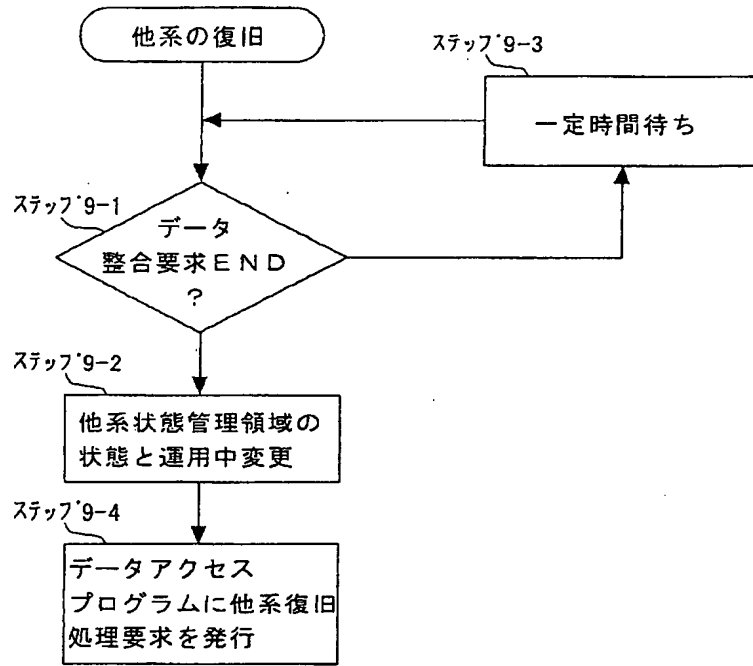
【図4】



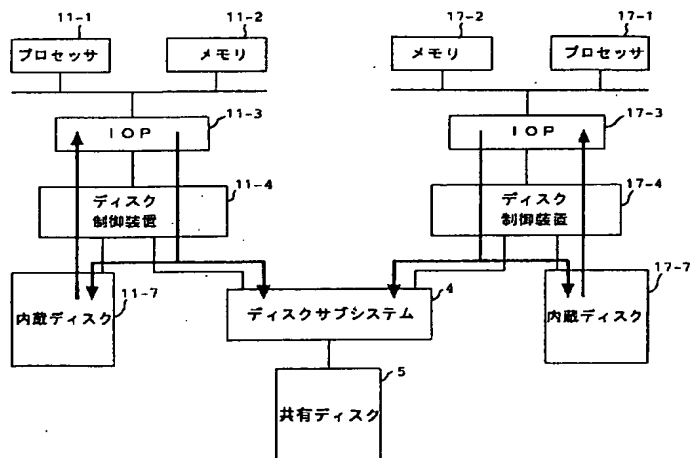
【図6】



【図9】



【図11】



【図10】

